Gweld y Gair: Standardisation and clinical application of a Welsh language measure for estimation of premorbid intellectual functioning

Clare, Linda; Toms, Gillian; Thomas, Enlli; Hindle, John; Woods, Robert

FPOP Bulletin

Published: 31/07/2017

Peer reviewed version

Gweld y Gair: standardisation and clinical application of a Welsh language measure for estimation of premorbid intellectual functioning

Linda Clare[1,2], Gill Toms[2], Enlli Thomas[2], John V Hindle[2] and Robert T Woods[2]

1. University of Exeter
2. Bangor University

Address for correspondence:
Professor Linda Clare
Centre for Research in Ageing and Cognitive Health (REACH)
School of Psychology
University of Exeter
Exeter EX4 4QG
United Kingdom
Email: l.clare@exeter.ac.uk
http://psychology.exeter.ac.uk/reach/

Abstract – English 199
Abstract – Welsh 219
Text and references 2527
Tables - 8

Abstract

Neuropsychological assessments, especially for suspected dementia, often emphasise the comparison of current and pre-morbid intellectual functioning. However, when assessing bilinguals, estimation of pre-morbid intellectual functioning may vary depending on which language provides the medium of testing. For bilingual Welsh-English speakers, testing in English only may yield an inaccurate profile of pre-morbid ability and hence affect the accuracy of subsequent diagnosis. We report the development and standardisation of a Welsh-language lexical decision task, Prawf Gweld y Gair, and evaluation of its suitability for assessing pre-morbid intellectual functioning in clinical groups. Standardisation with 101 healthy over 50s showed that scores correlated significantly with scores on other measures of crystallised intelligence and, when age and gender were taken into account, significantly predicted scores on a measure of non-verbal fluid intelligence. In subsequent evaluation with healthy older controls (n = 25), people who have Alzheimer's, vascular or mixed dementia (n = 26) and Parkinson's disease (n = 25), as predicted, there were no significant correlations with age, fluid intelligence or cognitive status in the clinical groups. Gweld y Gair shows promise as a Welsh-language test of premorbid intellectual functioning and may be useful for clinicians assessing the cognitive abilities of older Welsh speakers.

Mae asesiadau niwroseicolegol, yn arbennig pan yn amau dementia, yn aml yn rhoi pwyslais ar gymhariaeth o weithrediad deallusol cyfredol a chyn-forbid. Fodd bynnag, wrth asesu pobl ddwyieithog, gall amcangyfrif o weithrediad deallusol cyn-forbid amrywio yn dibynnu ar gyfrwng iaith y profion. Ar gyfer siaradwyr dwyieithog Cymraeg-Saesneg, gall profi yn Saesneg yn unig gynhyrchu proffil anghywir o allu cyn-forbid a thrwy hynny gall effeithio ar gywirdeb unrhyw ddiagnosis dilynol. Adroddwn yma ar ddatblygiad proses safoni tasg dewis geiriau Cymraeg, Prawf Gweld y Gair, a'n gwerthusiad o'i addasrwydd ar gyfer asesu gweithrediad deallusol cyn-forbid mewn grwpiau clinigol. Drwy safoni gyda 101 oedolyn iach dros 50 oed, dangosodd ein canlyniadau bod sgorau ar y prawf yn cydberthynnu'n arwyddocaol gyda sgorau ar fesurau eraill o wybodaeth cyn-forbid, a phan ystyriwyd oed a rhyw, roedd yn ragweld sgorau ar fesur o lifedd gwybodaeth di-eiriol. Mewn gwerthusiad dilynol gyda grŵp rheolaeth o bobl hŷn oedd yn iach (n = 25), pobl sydd â chlefyd Alzheimer, demensia fasgwlaidd neu gymysg (n = 26) a chlefyd Parkinson (n = 25), fel y rhagwelwyd, nid oedd unrhyw cydberthyniad arwyddocaol gydag oed, llifedd deallusrwydd na statws gwybyddol yn y grwpiau clinigol. Mae Gweld y Gair yn dangos addewid fel prawf Cymraeg o weithrediadau deallusol cyn-forbid a gall fod yn ddefnyddiol i glinigwyr asesu galluoedd gwybyddol siaradwyr Cymraeg hŷn.

***Gweld y Gair*: standardisation and clinical application of a Welsh language measure for estimation of premorbid intellectual functioning**

Clinical neuropsychology has paid relatively limited attention to issues of linguistic and cultural diversity. This is of particular concern to clinicians working with bilingual populations where such issues directly affect assessment, treatment and care. Clinicians may not have the necessary language skills to communicate with patients in their first or preferred language, standardised assessment measures may be unavailable in one or both languages, bilingual normative data is rarely available, and translated tests do not always have known, established psychometric qualities.

This is highly relevant to the situation of Welsh-English bilinguals in Wales. Since all Welsh-English bilinguals are typically fluent in English, clinicians often assume it appropriate to assess an individual's ability in English only. However, this approach might contribute to misdiagnosis. For example, Morgan and Crowder (2003) administered both a Welsh translation and the original English-language version of the Mini Mental State Examination (MMSE: Folstein, Folstein & McHugh, 1975) to 31 Welsh-English bilingual patients and found that limiting the testing of bilinguals to English alone would have underestimated the potential abilities of 42% of the sample. Therefore it would be useful to have the potential to assess ability through the medium of both English and Welsh.

Measures based on language ability are important in assessing suspected dementia because they provide clinically-useful estimates of premorbid IQ; for example the National Adult Reading Test (NART: Nelson, 1982) has long been considered a good

marker of intellectual and educational attainment across the lifespan, and is usually unimpaired in the early stages of dementia (Crawford, Deary, Starr & Whalley, 2001). More recently-developed measures that have adopted the same paradigm are the Wechsler Test of Adult Reading (WTAR) and Test of Premorbid Functioning – UK Version ($TOPF^{UK}$). These tests are aligned to versions of the Wechsler Adult Intelligence Scale (WAIS), allowing a direct comparison between premorbid and current functioning. Different reading tests, when compared, may produce different estimates of premorbid IQ (Norton, Watt, Gow & Crowe, 2016), and hence reading tests may not be the most accurate method for predicting premorbid IQ (Watt, Gow, Norton & Crowe, 2016). Furthermore, reading ability is a domain in which the potential to assess in both languages could be vital for accurate diagnosis. However, as these measures rely on irregular spelling-to-sound correspondence they do not lend themselves well to adaptation into Welsh, where most pronunciations are regular. Therefore a different approach is needed.

An alternative paradigm, feasible in Welsh, involves lexical decision tasks. The ability to discriminate between words and non-words is considered to be a reflection of crystallised knowledge, which may be relatively unimpaired in the early stages of dementia (Baddeley, Emslie & Nimmo-Smith, 1993). For example, the Spot-the-Word Test (STW; Baddeley, Emslie, & Nimmo-Smith, 1992) contains 60 word pairs of varying difficulty levels, each consisting of a word and an orthographically plausible non-word; the task is to identify the real word in each case. STW has good convergent and discriminative validity and the impact of dementia on performance is relatively minor in the early stages of the condition (McFarlane, Welch & Rogers, 2006).

We aimed to develop, standardise and conduct a preliminary evaluation of a Welsh lexical decision task based on the STW format, which presents participants with word pairs and asks them to distinguish the real word from the non-word.

Measure development

We began by identifying an initial pool of 120 Welsh words that range in difficulty based on their frequency in the language as noted in the Cronfa Electroneg o'r Gymraeg corpus of Welsh words (CEG: Ellis, O'Dochartaigh, Hicks, Morgan & Laporte, 2001). One-third of selected words were high frequency, one-third less frequent and one-third infrequent. For each selected word, we created a plausible non-word which was similar in length and number of syllables. The word pairs were reviewed by a Welsh language teacher and a Welsh language translator, and trialled with 20 Welsh speakers to assess acceptability.

We then administered this 120-item version of the task to 60 first-language Welsh-speakers who ranged in age from 20 to 81 years, recruited from among the University student population and from local clubs, groups and societies. We established test-retest reliability by asking 20 participants to complete the task again on a second occasion at least one week later. After assessing inter-item and test-retest correlations, we developed a final version of the task composed of 59 items. The Cronbach's alpha for the internal consistency of the scale items was .93 and test-retest reliability was .92. Item response consistency between initial test and re-test averaged 74.24%. We called the task 'Prawf Gweld y Gair' (GYG), which translates literally as 'See the

Word Test'. We undertook two studies to standardise and evaluate the new measure, which are reported in this paper.

Study One: Standardisation of GYG with healthy bilingual older adults

Method

We aimed to explore the correlations between GYG and other neuropsychological assessments, to compare scores achieved by participants with different levels of occupational and educational attainment, to derive percentile-based scores in a non-clinical sample, and to derive a regression formula for predicting a person's score on a test of fluid intelligence on the basis of the GYG score.

We recruited a convenience sample of participants aged 50 and over, drawn from local clubs, groups and societies, who completed GYG and other measures in a single assessment session. The assessment battery included an adult adaptation of the Welsh Vocabulary Test (WVT: Gathercole & Thomas, 2007), the Raven's Coloured Progressive Matrices (CPM: Raven, 1995), which provides a non-verbal estimate of fluid intelligence, and the English-language STW test (Baddeley et al., 1992). Ethical approval was granted by the appropriate University ethics committee.

We predicted that Gweld y Gair scores would correlate with scores on the WVT as both measures reflect 'crystallised' aspects of intelligence. However, we expected that the correlation with non-verbal fluid intelligence (as measured by the CPM) would be lower. Additionally, we predicted that participants with higher educational and

occupational attainment would score better on GYG, but GYG scores should not correlate significantly with age.

Results and discussion

The standardisation sample comprised 101 healthy individuals, 64% male, with a mean age of 65.66 years (range: 51 – 85). For those participants who provided information about their education, 66.3% had achieved educational qualifications. The sample was of relatively high socioeconomic status; 5.9% had held professional occupations, 57.4% had held managerial and technical occupations, 21.8% had held non-manual skilled occupations, 2% had held manual skilled occupations and 11.9% had held partly skilled occupations.

Table 1 contains a summary of the mean scores for GYG, STW, WVT and the CPM. We had complete data sets for 98 participants and the remaining 3 had incomplete responses to either GYG, STW or both. Table 2 shows the percentile scores for GYG and Table 3 details the correlations between GYG scores, age, socioeconomic status (SES) and scores on the other measures.

(((Tables 1, 2 and 3 near here)))

Age did not correlate with scores on any measure. GYG scores correlated significantly with STW score, suggesting that ability in completing a lexical decision task in one language relates to ability to perform the task in the second language. Both GYG and STW scores were significantly correlated with scores on the WVT. This suggests that all three tests were tapping crystallised intelligence, as predicted. Only

the WVT correlated significantly with scores on the CPM. Educational level is likely to be a strong moderator of performance on vocabulary tests and may influence performance on fluid intelligence tests. As predicted, GYG and STW scores correlated less strongly with scores on the CPM, and these correlations were not statistically significant. SES was negatively correlated with scores on GYG and STW, suggesting that individuals with higher SES performed better on these lexical decision tasks, as shown in Table 4.

(((Table 4 near here)))

CPM score was significantly predicted by a combination of GYG score, age and gender ($R=0.379$, adjusted $R^2 = .113$, $F=4.65$, $p=0.005$). The regression formula for predicting CPM score from GYG score was: Predicted CPM total = 28.15 + (.099*GYG) – (.09*Age) + (2.06*Gender) + (0.147*Age left school), where gender is scored 1 = male and 2 = female, and Age left school is age in years at the time of leaving school.

Comparing GYG and STW scores, after correcting for the different number of items, 68 participants scored better on GYG than STW, 29 scored better on STW than GYG, and one participant scored the same on both measures. While there was no expectation that the two measures should be exactly equal in difficulty level, it was nevertheless of interest to explore these differences further. Difference scores (GYG minus STW) were calculated for all participants. These ranged from -16 (STW better) to +21 (GYG better), with a mean of 2.59 (SD 8.42). We subdivided the participants into three groups based on the difference scores. Group one participants had a

difference score within one standard deviation of the difference score mean (GYG=STM, n = 70; difference score of between -5.83 and +11.01). Group two participants had a difference score at least one standard deviation above the mean with GYG scores better than STW scores (GYG>STW, n = 12; difference score of +12 and above). Participants in the third group had a difference score at least one standard deviation below the mean with STW scores better than GYG scores (STW>GYG, n = 16; difference score of -6 and below). Table 5 shows the mean scores on all measures for each subgroup. Comparison of scores for each subgroup using one-way ANOVA revealed no significant between-group differences in age or scores on the WVT or CPM.

(((Table 5 near here)))

Study Two: Evaluating GYG in an older clinical sample

Method

We administered GYG to three groups: people with Alzheimer's or mixed Alzheimer's/ vascular dementia, people with Parkinson's disease, and healthy age-matched controls. We recruited people with Alzheimer's disease and Parkinson's disease from local National Health Service (NHS) memory clinics and movement disorders clinics. We recruited healthy older people through newspaper articles and presentations to community groups. The study was approved by the relevant University and NHS ethics committees. People with dementia were in the early stages, indicated by a MMSE score of 18 or above, and able to give informed consent

for participation. All participants completed a short test battery including GYG, STW, WVT, either CPM or the Raven's Standard Progressive Matrices (SPM; Raven et al., 1998), and the NART. We used the NART rather than its more recent counterparts because it was used in the original development of STW (Baddeley et al., 1992) and was therefore appropriate for purposes of comparison.

Results and discussion

Table 6 presents the demographic and, where relevant, clinical characteristics for each participant group. For the participants with Parkinson's disease, 76% were at Hoehn and Yahr (1967) stage 1, 16% were at stage 2 and 8% were at stage three.

(((Table 6 near here)))

As is evident in Table 7, participants with Alzheimer's or mixed Alzheimer's/ vascular dementia performed similarly to participants with Parkinson's Disease on GYG and STW, with both groups performing more poorly than control participants. The mean score equated to a percentile rank of 16.2 for people with dementia and 14.1 for people with Parkinson's. The control group mean equated to a percentile rank of 40.4.

(((Table 7 near here)))

Table 8 presents the correlation between GYG scores and performance on the other measures according to participant group. The correlations for the control group were similar to those in the standardisation study, apart from the lack of a significant

correlation with SES, which may be due to the smaller sample size in this clinical study. In addition, GYG score was significantly correlated with NART score.
For participants with Parkinson's, GYG score was correlated only with performance on the WVT, while for participants with dementia, GYG score did not correlate with any other variables of interest. It is promising, that, as predicted, GYG did not correlate with age, fluid intelligence (CPM), or cognitive status (MMSE score) in the clinical groups.

(((Table 8 near here)))

The lack of association between STW and GYG highlights the potential for language to influence test performance. Participants did not always find it easier to complete the lexical decision task in Welsh. It could be that the final GYG version was somewhat challenging for adults with cognitive or neurological impairment, especially where levels of education were relatively low. Another explanation could be that the cohort studied, while fluent Welsh speakers, were less confident in reading Welsh. Most participants in this age-group will have been educated through the medium of English due to government policy at that time, and might be more confident reading English.

Overall discussion

This study represents a novel step in developing neuropsychological assessment measures that are linguistically and culturally appropriate for Welsh speakers. GYG is a lexical decision task suitable for use in Welsh populations which provides an estimate of crystallized (or premorbid) intelligence and has satisfactory internal

consistency, test retest reliability and concurrent validity. While the task would ideally be administered in the context of an assessment conducted through the medium of Welsh, GYG can be administered by clinicians who are not Welsh speakers as part of an assessment conducted in English. GYG shows some promise as a test of premorbid IQ.

People with dementia as well as participants with Parkinson's disease performed less well on GYG compared to healthy controls. However, the control group participants had higher levels of education, and this may partly account for the differences observed, although it has been noted that most tests of crystallized ability are vulnerable to some decline in the presence of neurological disease (Crawford et al., 2001). Unlike STW (Baddeley et al., 1992), age does not significantly influence GYG performance, although SES does affect scores on both GyG and STW. The test may prove to be more salient for the next generation of older adults who are more likely to have been educated through the medium of Welsh and may be more confident reading Welsh. The next step would be to test the properties of GYG in a larger clinical sample.

Welsh is the preferred language for many residents of Wales, and our findings highlight the need for more Welsh language tests and the necessity of establishing bilingual norms to ensure accurate assessment. The Gweld y Gair lexical decision task represents one step in this direction and may be useful for clinicians working with Welsh speakers.

Acknowledgements

We gratefully acknowledge the assistance with data collection provided by Delyth Jones, Pamela Martin-Forbes, Lexi Hindley, Siwan Roberts and Eleri Jones.

The standardisation study was funded by grants from the North West Wales NHS Trust Research and Development Committee to L. Clare, E. Thomas, and R.T. Woods and from the North Wales Research Committee to E. Thomas, L. Clare and R.T. Woods. Data for the clinical study was collected as part of the Bilingualism and Neurodegenerative Conditions (BANC) project, which was funded by Economic and Social Research Council grant RES-062-23-1931 to L. Clare, J.V. Hindle, V.C. Gathercole, E. Thomas, E. Bialystok, F. Craik, and C. Whitaker. Further details of Prawf Gweld y Gair may be obtained from the first author.

References

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Spot the Word test.* Bury St Edmunds, UK: Thames Valley Test Company.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot the Word test: a robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology, 32 (Pt 1),* 55-65. doi: 10.1111/j.2044-8260.1993.tb01027.x

Crawford, J. R., Deary, I. J., Starr, J., & Whalley, L. J. (2001). The NART as an index of prior intellectual functioning: a retrospective validation study covering a 66-year interval. *Psychological Medicine, 31 (3),* 451-458. doi: 10.1017/S0033291701003634

Ellis, N. C., O'Dochartaigh, C., Hicks, W., Morgan, M., & Laporte, N. (2001). *Cronfa Electroneg O Gymraeg (CEG): a 1 million lexical database and frequency count for welsh.* Bangor, UK: University of Wales [Released on the internet, July]

Folstein, M. F., Folstein, F. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12 (3), 189-198. doi: 10.1016/0022-3956(75)900265-6

Gathercole, V. C. M., & Thomas, E. M. (2007). *Prawf Geirfa Cymraeg*. Retrieved from: www.pgc.bangor.ac.uk

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression and mortality. *Neurology, 17 (5),* 427-442. doi: 10.1212/wnl.17.5.427

McFarlane, J., Welch, J., & Rogers, J. (2006). Severity of Alzheimer's disease and effect on premorbid measures of intelligence. British Journal of Clinical Psychology, 45 (part 4), 453-63. doi: 10.1348/014466505X71245

Morgan, T., & Crowder, R. (2003). Mini mental state examinations in English: Are they suitable for people with dementia who are Welsh speaking. *Dementia, 2 (2),* 267-272. doi: 10.1177/1471301203002002009

Nelson, H. E. (1982). National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual. Berkshire, UK: NFER-Nelson.

Norton, K., Watt, S., Gow, B. and Crowe, S. F. (2016). Are Tests of Premorbid Functioning Subject to the Flynn Effect? *Australian Psychologist, 51,* 374–379. doi:10.1111/ap.12235

Raven, J. C. (1995). *Manual for the coloured progressive matrices (revised).* Windsor, UK: NFRE-Nelson.

Raven, J., Raven, J. C. & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales.* Oxford: Oxford Psychologists Press.

Watt, S., Gow, B., Norton, K. and Crowe, S. F. (2016). Investigating Discrepancies between Predicted and Observed Wechsler Adult Intelligence Scale-Version IV Full-Scale Intelligence Quotient Scores in a Non-Clinical Sample. *Australian Psychologist, 51,* 380–388. doi:10.1111/ap.12239

Table 1. Mean scores on all measures for the standardisation sample

Measure *(Maximum score)*	**Number in analysis**	**Mean**	**Standard deviation**	**Range**
Gweld y Gair *(59)*	99	46.55	9.03	21-59
Spot the Word *(60)*	98	43.87	7.50	25-58
Welsh Vocabulary Test *(102)*	99	93.44	5.31	75-99
Ravens Coloured Matrices *(36)*	99	33.20	3.17	20-36

Table 2. Percentile scores for Gweld y Gair in the standardisation sample (n = 99 healthy older controls)

Gweld y Gair Score *(max 59)*	**Frequency in standardisation sample**	**Cumulative Percentile**
21	1	1
22	1	2
28	1	3
29	1	4
30	1	5
31	3	8
32	3	11
33	1	12
34	2	14
35	2	16
37	1	17
38	1	18
39	2	20
40	3	23
41	4	27
42	5	32
43	3	35
44	1	36
45	4	40
46	3	43
47	2	46
48	2	48
49	3	51
50	7	58
51	2	60
52	3	63
53	12	75
54	5	80
55	5	85
56	5	90
57	7	97
58	2	99
59	1	100

Table 3. Correlations between scores on all measures, socioeconomic status and age in the standardisation sample (n = 101 healthy older controls)

	GYG	**STW**	**WVT**	**CPM**	**Age**	**SES[$]**	**Qual[$]**	**School**
GYG		.495** *(N: 98)*	.474** *(N:99)*	.193 *(N:99)*	.075 *(N:99)*	-.306** *(N:99)*	-.221 *(N:72)*	.366** *(N:87)*
STW	.495** *(N:98)*		.450** *(N:99)*	.130 *(N:99)*	-.124 *(N:99)*	-.297** *(N:99)*	.209 *(N:71)*	.366** *(N:87)*
WVT	.474** *(N:99)*	.450** *(N:99)*		.275** *(N:101)*	-.106 *(N:101)*	-.103 *(N:101)*	.023 *(N:72)*	.090 *(N:87)*
CPM	.193 *(N:99)*	.130 *(N:99)*	.275** *(N:101)*		-.148 *(N:101)*	.077 *(N:101)*	-.032 *(N:72)*	-.147 *(N:87)*
Age	.075 *(N:99)*	-.124 *(N:99)*	-.106 *(N:101)*	-.148 *(N:101)*		.131 *(N:101)*	.032 *(N:73)*	-.128 *(N:89)*
SES[$]	-.306** *(N:99)*	-.297** *(N:99)*	-.103 *(N:101)*	.077 *(N:101)*	.131 *(N:101)*		-.459** *(N:73)*	-.472** *(N:89)*
Qual[$]	-.221 *(N:72)*	.209 *(N:71)*	.023 *(N:72)*	-.032 *(N:72)*	.032 *(N:73)*	-.459** *(N:73)*		.429** *(N:63)*
School	.366** *(N:87)*	.366** *(N:87)*	.090 *(N:87)*	-.147 *(N:87)*	-.128 *(N:89)*	-.472** *(N:89)*	.429** *(N:63)*	

Abbreviations: GYG- Gweld y Gair, STW- Spot the Word, WVT- Welsh Vocabulary Test, CPM- Coloured Progressive Matrices, SES- Socioeconomic Status, Qual- Qualifications achieved, School- Age left school

*Pearson correlation: ** = significant at the .01 level (two-tailed)*
*[$]Spearman Rho correlation: ** = significant at the .01 level (two-tailed)*

Table 4. Comparison of mean scores across different levels of socioeconomic status in the standardisation sample

	Professional (n= 6)	**Managerial & Technical (n = 58)**	**Skilled, Non-manual (n = 22)**	**Skilled, Manual (n = 2)**	**Partly Skilled (n = 12)**
GYG	50.67	48.04	43.14	46.00	42.75
STW	48.00	45.40	41.00	51.00	39.75
WVT	96.17	93.14	94.68	94.00	90.25
CPM	32.50	33.31	33.64	29.50	33.00

Abbreviations: GYG- Gweld y Gair, STW- Spot the Word, WVT- Welsh Vocabulary Test, CPM- Coloured Progressive Matrices

Table 5. Subgroups based on difference scores between GYG and STW scores within the standardisation sample (n = 98 healthy older controls)

	GYG=STW	**GYG>STW**	**STW>GYG**
Number of participants	70	12	16
Gender	60 M: 10 F	1 M: 11 F	2 M: 14 F
Mean age	64.87	70.25	64.81
GYG mean	48.24	51.58	34.81
STW mean	44.66	36.42	46.00
WVT mean	93.59	95.50	91.63
CPM mean	33.09	34.92	32.69

Abbreviations: GYG- Gweld y Gair, STW- Spot the Word, WVT- Welsh Vocabulary Test, CPM- Coloured Progressive Matrices, M- Male, F- Female

Table 6. Demographic details for participants in the clinical study

	Participants with dementia	**Participants with Parkinson's disease**	**Control participants**
Number of participants	26	25	25
Gender (Male: Female)	14:12	17:8	13:12
Mean age (Range)	79.85 (69-87)	68.92 (40-85)	71.64 (62-97)
Educational qualifications achieved	11%	44%	76%
Mean age (SD) school leaving	15.68 (1.76)	15.84 (.99)	16.80 (1.41)
Professional occupations held	11.5%	4%	12%
Managerial and technical occupations held	34.6%	32%	52%
Skilled (manual/ non-manual) jobs held	3.8%/ 19.2%	20%/ 16%	12%/ 12%
Partly skilled jobs held	11.5%	20%	8%
Unskilled jobs held	0%	4%	0%
Mean MMSE (SD)	22.72 (3.20)	28.24 (1.78)	28.92 (1.41)

Table 7: Means, standard deviations and ranges for all tests according to participant group in the clinical study

	GYG *Max score: 59*	**STW** *Max score: 60*	**WVT** *Max score:56*	**CPM** *Max score:36*	**SPM** *Max score:60*	**NART** *Max score:50*
AD or mixed dementia	35.04 (7.77) R:13-54 *N:26*	44.32 (5.51) R:30-55 *N:22*	43.05 (9.04) R:12-54 *N:22*	23.33 (8.09) R:2-33 *N:24*		25.86 (10.61) R:7-40 *N:21*
PD	34.32 (7.31) R:23-50 *N:25*	47.312 (6.26) R:32-57 *N:24*	48,00 (7.40) R:29-56 *N:24*	30.08 (4.58) R:23-36 *N:25*		27.88 (7.43) R:11-41 *N:24*
Controls	45.60 (10.57) R:24-58 *N:25*	50.05 (5.68) R:41-58 *N:21*	53.17 (3.89) R:39-56 *N:24*		40.25 (9.17) R:19-58 *N:24*	36.00 (8.50) R:16-47 *N:25*

Code: GYG=Gweld y Gair, STW=Spot the Word, WVT=Welsh Vocabulary Test, NART=National Adult Reading Test, CPM=Coloured Progressive Matrices, SPM=Standard Progressive Matrices, ( ) =Standard Deviation, R=Range, N=Number of participants in analysis

Table 8: Correlations between Gweld y Gair and other measures in the clinical study groups

	STW	**WVT**	**NART**	**CPM/ SPM**	**MMSE**	**Age**	**SES[$]**
GYG Alzheimer's/ Mixed	.117 *N:22*	-.053 *N:23*	.121 *N:21*	-.166 *N:24*	.369 *N:25*	.047 *N:26*	-.318 *N:21*
GYG Parkinson's Disease	.115 *N:24*	.621** *N:24*	.183 *N:24*	-.059 *N:25*	.064 *N:25*	.360 *N:25*	-.143 *N:24*
GYG Control participants	.498* *N:21*	.661** *N:24*	.572** *N:25*	.163 *N:24*	.283 *N:25*	.303 *N:25*	-.064 *N:24*

Code: GYG=Gweld y Gair, STW=Spot the Word, WVT=Welsh Vocabulary Test, NART=National Adult Reading Test, CPM/SPM=Coloured Progressive Matrices/ Standard Progressive Matrices, MMSE=Mini-Mental State Examination, SES=Socioeconomic status, N=Number of participants in analysis

Note: Pearson's correlations except [$]Spearman's Rho correlation, *significant at 0.05 level